

Durham Research Online

Deposited in DRO:

14 May 2009

Version of attached file:

Accepted Version

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Troffaes, M. C. M. (2007) 'Decision making under uncertainty using imprecise probabilities.', International journal of approximate reasoning., 45 (1). pp. 17-29.

Further information on publisher's website:

<http://dx.doi.org/10.1016/j.ijar.2006.06.001>

Publisher's copyright statement:

Additional information:

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

Decision Making under Uncertainty using Imprecise Probabilities

Matthias C. M. Troffaes

*Carnegie Mellon University, Department of Philosophy, Baker Hall 135,
Pittsburgh, PA 15213, US*

Abstract

Various ways for decision making with imprecise probabilities—admissibility, maximal expected utility, maximality, E-admissibility, Γ -maximax, Γ -maximin, all of which are well-known from the literature—are discussed and compared. We generalize a well-known sufficient condition for existence of optimal decisions. A simple numerical example shows how these criteria can work in practice, and demonstrates their differences. Finally, we suggest an efficient approach to calculate optimal decisions under these decision criteria.

Key words: decision, optimality, uncertainty, probability, maximality, E-admissibility, maximin, lower prevision

1 Introduction

Often, we find ourselves in a situation where we have to make some decision d , which we may freely choose from a set D of available decisions. Usually, we do not choose d arbitrarily in D : indeed, we wish to make a decision that performs best according to some criterion, i.e., an *optimal* decision. It is commonly assumed that each decision d induces a real-valued gain J_d : in that case, a decision d is considered optimal in D if it induces the highest gain among all decisions in D . This holds for instance if each decision induces a lottery over some set of rewards, and these lotteries form an ordered set satisfying the axioms of von Neumann Morgenstern

Email address: `matthias.troffaes@gmail.com` (Matthias C. M. Troffaes).

[1], or more generally, the axioms of for instance Herstein and Milnor [2], if we wish to account for unbounded gain.

So, we wish to identify the set $\text{opt}(D)$ of all decisions that induce the highest gain. Since, at this stage, there is no uncertainty regarding the gains J_d , $d \in D$, the solution is simply

$$\text{opt}(D) = \arg \max_{d \in D} J_d. \quad (1)$$

Of course, $\text{opt}(D)$ may be empty; however, if the set $\{J_d: d \in D\}$ is a compact subset of \mathbb{R} —this holds for instance if D is finite—then $\text{opt}(D)$ contains at least one element. Secondly, note that even if $\text{opt}(D)$ contains more than one decision, all decisions d in $\text{opt}(D)$ induce the same gain J_d ; so, if, in the end, the gain is all that matters, it suffices to identify only one decision d^* in $\text{opt}(D)$ —often, this greatly simplifies the analysis.

However, in many situations, the gains J_d induced by decisions d in D are influenced by variables whose values are uncertain. Assuming that these variables can be modelled through a random variable X that takes values in some set \mathcal{X} (the *possibility space*), it is customary to consider the gain J_d as a so-called *gamble* on X , that is, we view J_d as a real-valued gain that is a bounded function of X , and that is expressed in a fixed state-independent utility scale. So, J_d is a bounded \mathcal{X} – \mathbb{R} -mapping, interpreted as an uncertain gain: taking decision d , we receive an amount of utility $J_d(x)$ when x turns out to be the realisation of X . For the sake of simplicity, we shall assume that the outcome x of X is independent of the decision d we take: this is called *act-state independence*. What decision should we take?

Irrespective of our beliefs about X , a decision d in D is not optimal if its gain gamble J_d is *point-wise dominated* by a gain gamble J_e for some e in D , i.e., if there is an e in D such that $J_e(x) \geq J_d(x)$ for all $x \in \mathcal{X}$ and $J_e(x) > J_d(x)$ for at least one $x \in \mathcal{X}$: choosing e guarantees a higher gain than choosing d , possibly strictly higher, regardless of the realisation of X . So, as a first selection, let us remove all decisions from D whose gain gambles are point-wise dominated (see Berger [3, Section 1.3.2, Definition 5 ff., p. 10]):

$$\text{opt}_{\geq}(D) := \{d \in D: (\forall e \in D)(J_e \not\geq J_d \text{ or } J_e = J_d)\} \quad (2)$$

where $J_e \geq J_d$ is understood to be point-wise, and $J_e \not\geq J_d$ is understood to be the negation of $J_e \geq J_d$. The decisions in $\text{opt}_{\geq}(D)$ are called *admissible*, the other decisions in D are called *inadmissible*. Note that we already recover Eq. (1) if there is no uncertainty regarding the gains J_d , i.e., if all J_d are constant functions of X . When do admissible decisions exist? The set $\text{opt}_{\geq}(D)$ is non-empty if $\{J_d: d \in D\}$ is a non-empty and weakly compact subset of the set $\mathcal{L}(\mathcal{X})$ of all gambles on \mathcal{X} (see

Theorem 3 further on). Note that this condition is sufficient, but not necessary.

In what follows, we shall try to answer the following question: given additional information about X , how can we further reduce the set $\text{opt}_{\geq}(D)$ of admissible decisions? The paper is structured as follows. Section 2 discusses the classical approach of maximising expected utility, and explains why it is not always a desirable criterion for selecting optimal decisions. Those problems are addressed in Section 3, discussing alternative approaches to deal with uncertainty and optimality, all of which attempt to overcome the issues raised in Section 2, and all of which are known from the literature. Finally, Section 4 compares these alternative approaches, and explains how optimal decisions can be obtained in a computationally efficient way. A few technical results are deferred to the appendix, where we, among other things, generalize a well-known technical condition on the existence of optimal decisions.

2 Maximising Expected Utility?

In practice, beliefs about X are often modelled by a (possibly finitely additive) probability measure μ on a field \mathcal{F} of subsets of \mathcal{X} , and one then arrives at a set of optimal decisions by maximising their expected utility with respect to μ ; see for instance Raiffa and Schlaifer [4, Section 1.1.4, p. 6], Levi [5, Section 4.8, p. 96, ll. 23–26], or Berger [3, Section 1.5.2, Paragraph I, p. 17]. Assuming that the field \mathcal{F} is sufficiently large such that the gains J_d are measurable with respect to \mathcal{F} —this means that every J_d is a uniform limit of \mathcal{F} -simple gambles—the *expected utility* of the gain gambles J_d is given by:

$$\mathbf{E}_{\mu}(J_d) := \int J_d d\mu,$$

where we take for instance the Dunford integral on the right hand side; see Dunford [6, p. 443, Sect. 3], and Dunford and Schwartz [7, Part I, Chapter III, Definition 2.17, p. 112]—this linear integral extends the usual textbook integral (see for instance Kallenberg [8, Chapter 1]) to case where μ is not σ -additive. Recall that we have assumed act-state independence: μ is independent of d .

As far as it makes sense to rank decisions according to the expected utility of their gain gambles, we should *maximise expected utility*:

$$\text{opt}_{\mathbf{E}_{\mu}}(D) := \arg \max_{d \in \text{opt}_{\geq}(D)} \mathbf{E}_{\mu}(J_d). \quad (3)$$

When do optimal solutions exist? The set $\text{opt}_{\mathbf{E}_\mu}(D)$ is guaranteed to be non-empty if $\{J_d: d \in D\}$ is a non-empty and compact subset of the set $\mathcal{L}(\mathcal{X})$ of all gambles on \mathcal{X} , with respect to the supremum norm. Actually, this technical condition is sufficient for existence with regard to all of the optimality conditions we shall discuss further on. Therefore, without further ado, we shall assume that $\{J_d: d \in D\}$ is non-empty and compact with respect to the supremum norm. A slightly weaker condition is assumed in Theorem 5, in the appendix of this paper.

Unfortunately, it may happen that our beliefs about X cannot be modelled by a probability measure, simply because we have insufficient information to identify the probability $\mu(A)$ of every event A in \mathcal{F} . In such a situation, maximising expected utility usually fails to give an adequate representation of optimality.

For example, let X be the unknown outcome of the tossing of a coin; say we *only* know that the outcome will be either heads or tails (so $\mathcal{X} = \{H, T\}$), and that the probability of heads lays between 28% and 70%. Consider the decision set $D = \{1, 2, 3, 4, 5, 6\}$ and the gain gambles

$$\begin{array}{ll} J_1(H) = 4, & J_1(T) = 0, \\ J_2(H) = 0, & J_2(T) = 4, \\ J_3(H) = 3, & J_3(T) = 2, \\ J_4(H) = \frac{1}{2}, & J_4(T) = 3, \\ J_5(H) = \frac{47}{20}, & J_5(T) = \frac{47}{20}, \\ J_6(H) = \frac{41}{10}, & J_6(T) = -\frac{3}{10}, \end{array}$$

Clearly, $\text{opt}_{\geq}(D) = \{1, 2, 3, 4, 5, 6\}$, and

$$\text{opt}_{\mathbf{E}_\mu}(D) = \begin{cases} \{2\}, & \text{if } \mu(H) < \frac{2}{5}, \\ \{2, 3\}, & \text{if } \mu(H) = \frac{2}{5}, \\ \{3\}, & \text{if } \frac{2}{5} < \mu(H) < \frac{2}{3}, \\ \{1, 3\}, & \text{if } \mu(H) = \frac{2}{3}, \\ \{1\}, & \text{if } \mu(H) > \frac{2}{3}. \end{cases}$$

Concluding, if we have *no* additional information about X , but still insist on using a particular (and necessarily arbitrary) μ , which is only required to satisfy $0.28 \leq \mu(H) \leq 0.7$, we find that $\text{opt}_{\mathbf{E}_\mu}(D)$ is not very robust against changes in μ . This shows that maximising expected utility fails to give an adequate representation of optimality in case of ignorance about the precise value of μ .

3 Generalising to Imprecise Probabilities

Of course, if we have sufficient information such that μ can be identified, nothing is wrong with Eq. (3). We shall therefore try to generalise Eq. (3). In doing so, following Walley [9], we shall assume that our beliefs about X are modelled by a real-valued mapping \underline{P} defined on a—possibly only very small—set \mathcal{K} of gambles, that represents our assessment of the *lower expected utility* $\underline{P}(f)$ for each gamble f in \mathcal{K} ;¹ note that \mathcal{K} can be chosen empty if we are completely ignorant. Essentially, this means that instead of a single probability measure on \mathcal{F} , we now identify a closed convex set \mathcal{M} of finitely additive probability measures μ on \mathcal{F} , described by the linear inequalities

$$(\forall f \in \mathcal{K})(\underline{P}(f) \leq \mathbf{E}_\mu(f)). \quad (4)$$

We choose the domain \mathcal{F} of the measures μ sufficiently large such that all gambles of interest, in particular those in \mathcal{K} and the gain gambles J_d , are measurable with respect to \mathcal{F} . Without loss of generality, we can assume \mathcal{F} to be the power set of \mathcal{X} , although in practice, it may be more convenient to choose a smaller field.

For a given \mathcal{F} -measurable gamble g , not necessarily in \mathcal{K} , we may also derive a lower expected utility $\underline{\mathbf{E}}_{\underline{P}}(g)$ by minimising $\mathbf{E}_\mu(g)$ subject to the above constraints, and an upper expected utility $\overline{\mathbf{E}}_{\underline{P}}(g) = -\underline{\mathbf{E}}_{\underline{P}}(-g)$ by maximising $\mathbf{E}_\mu(g)$ over the above constraints. In case \mathcal{X} and \mathcal{K} are finite, this simply amounts to solving a linear program.

In the literature, \mathcal{M} is called a *credal set* (see for instance Giron and Rios [10], and Levi [5, Section 4.2, pp. 76–78], for more comments on this model), and \underline{P} is called a *lower prevision* (because they generalise the previsions, which are fair prices, of De Finetti [11, Vol. I, Section 3.1, pp. 69–75]).

The mapping $\underline{\mathbf{E}}_{\underline{P}}$ obtained, corresponds exactly to the so-called *natural extension* of \underline{P} (to the set of \mathcal{F} -measurable gambles), where $\underline{P}(f)$ is interpreted as a supremum buying price for f (see Walley [9, Section 3.4.1, p. 136]). In this interpretation, for any $s < \underline{P}(f)$, we are willing to pay any utility $s < \underline{P}(f)$ prior to observation of X , if we are guaranteed to receive $f(x)$ once x turns out to be the outcome of X . The natural extension then corresponds to the highest price we can obtain for an arbitrary gamble g , taken into account the assessed prices $\underline{P}(f)$ for

¹ The upper expected utility of a gamble f is $\overline{P}(f)$ if and only if the lower expected utility of $-f$ is $-\overline{P}(f)$. So, for any gamble f in \mathcal{K} , $\underline{P}(-f) = -\overline{P}(f)$, and therefore, without loss of generality, we can restrict ourselves to lower expected utility.

$f \in \mathcal{K}$. Specifically,

$$\underline{\mathbf{E}}_{\underline{P}}(g) = \sup \left\{ \alpha + \sum_{i=1}^n \lambda_i \underline{P}(f_i) : \alpha + \sum_{i=1}^n \lambda_i f_i \leq g \right\}, \quad (5)$$

where α varies over \mathbb{R} , n over \mathbb{N} , $\lambda_1, \dots, \lambda_n$ vary over \mathbb{R}^+ , and f_1, \dots, f_n over \mathcal{K} .

It may happen that \mathcal{M} is empty, in which case $\underline{\mathbf{E}}_{\underline{P}}$ is undefined (the supremum in Eq. (5) will always be $+\infty$). This occurs exactly when \underline{P} incurs a *sure loss* as a lower prevision, that is, if we can find a finite collection of gambles f_1, \dots, f_n in \mathcal{K} such that $\sum_{i=1}^n \underline{P}(f_i) > \sup [\sum_{i=1}^n f_i]$, which means that we are willing to pay more for this collection than we can ever gain from it, which makes no sense of course.

Finally, it may happen that $\underline{\mathbf{E}}_{\underline{P}}$ does not coincide with \underline{P} on \mathcal{K} . This points to a form of *incoherence* in \underline{P} : this situation occurs exactly when we can find a finite collection of gambles f_0, f_1, \dots, f_n and non-negative real numbers $\lambda_1, \dots, \lambda_n$, such that

$$\alpha + \sum_{i=1}^n \lambda_i f_i \leq f_0, \text{ but also } \underline{P}(f_0) < \alpha + \sum_{i=1}^n \lambda_i \underline{P}(f_i).$$

This means that we can construct a price for f_0 , using the assessed prices $\underline{P}(f_i)$ for f_i , which is strictly higher than $\underline{P}(f_0)$. In this sense, $\underline{\mathbf{E}}_{\underline{P}}$ corrects \underline{P} , as is apparent from Eq. (5).

Although the belief model described above is not the most general we may think of, it is sufficiently general to model both expected utility and complete ignorance: these two extremes are obtained by taking \mathcal{M} either equal to a singleton, or equal to the set of all finitely additive probability measures on \mathcal{F} (i.e., $\mathcal{K} = \emptyset$). It also allows us to demonstrate the differences between different ways to make decisions with imprecise probabilities on the example we presented before.

In that example, the given information can be modelled by, say, a lower prevision \underline{P} on $\mathcal{K} = \{I_H, -I_H\}$, defined by $\underline{P}(I_H) = 0.28$ and $\underline{P}(-I_H) = -0.7$, where I_H is the gamble defined by $I_H(H) = 1$ and $I_H(T) = 0$. For this \underline{P} , the set \mathcal{M} corresponds exactly to the set of all probability measures μ on $\mathcal{F} = \{\emptyset, \{H\}, \{T\}, \{H, T\}\}$, such that $0.28 \leq \mu(H) \leq 0.7$. We also easily find for any gamble f on X that

$$\underline{\mathbf{E}}_{\underline{P}}(f) = \min\{0.28f(H) + 0.72f(T), 0.7f(H) + 0.3f(T)\}.$$

3.1 Γ -Maximin and Γ -Maximax

As a very simple way to generalise Eq. (3), we could take the lower expected utility $\underline{\mathbf{E}}_P$ as a replacement for the expected utility \mathbf{E}_μ (see for instance Gilboa and Schmeidler [12], or Berger [3, Section 4.7.6, pp. 215–223]):

$$\text{opt}_{\underline{\mathbf{E}}_P}(D) := \arg \max_{d \in \text{opt}_{\geq}(D)} \underline{\mathbf{E}}_P(J_d); \quad (6)$$

this criterion is called Γ -*maximin*, and amounts to worst-case optimisation: we take a decision that maximises the worst expected gain. For example, if we consider the decision as a game against nature, who is assumed to choose a distribution in \mathcal{M} aimed at minimizing our expected gain, then the Γ -maximin solution is the best we can do. Applied on the example of Section 2, we find as a solution $\text{opt}_{\underline{\mathbf{E}}_P}(D) = \{5\}$.

In case $\mathcal{K} = \emptyset$, i.e., in case of complete ignorance about X , it holds that $\underline{\mathbf{E}}_P(f) = \inf_{x \in \mathcal{X}} f(x)$. Hence, in that case, Γ -maximin coincides with maximin (see Berger [3, Eq. (4.96), p. 216]), ranking decisions by the minimal (or infimum, to be more precise) value of their gain gambles.

Some authors consider best-case optimisation, taking a decision that maximises the best expected gain (see for instance Satia and Lave [13]). In our example, the “ Γ -maximax” solution is $\text{opt}_{\overline{\mathbf{E}}_P}(D) = \{2\}$.

3.2 Maximality

Eq. (3) is essentially the result of pair-wise preferences based on expected utility: defining the strict partial order $>_\mu$ on D as $d >_\mu e$ whenever $\mathbf{E}_\mu(J_d) > \mathbf{E}_\mu(J_e)$, or equivalently, whenever $\mathbf{E}_\mu(J_d - J_e) > 0$, we can simply write

$$\text{opt}_{\mathbf{E}_\mu}(D) = \max_{>_\mu} \left(\text{opt}_{\geq}(D) \right),$$

where the operator $\max_{>_\mu}(\cdot)$ selects the $>_\mu$ -maximal, i.e., the $>_\mu$ -undominated elements from a set with strict partial order $>_\mu$.

Using the supremum buying price interpretation, it is easy to derive pair-wise preferences from \underline{P} : define $>_{\underline{P}}$ as $d >_{\underline{P}} e$ whenever $\underline{\mathbf{E}}_P(J_d - J_e) > 0$. Indeed, $\underline{\mathbf{E}}_P(J_d - J_e) > 0$ means that we are disposed to pay a strictly positive price in order to take decision d instead of e , which clearly indicates strict preference of

d over e (see Walley [9, Sections 3.9.1–3.9.3, pp. 160–162]). Since $>_{\underline{P}}$ is a strict partial order, we arrive at

$$\begin{aligned}\text{opt}_{>_{\underline{P}}}(D) &:= \max_{>_{\underline{P}}}(\text{opt}_{\geq}(D)) \\ &= \{d \in \text{opt}_{\geq}(D) : (\forall e \in \text{opt}_{\geq}(D))(\underline{\mathbf{E}}_{\underline{P}}(J_e - J_d) \leq 0)\}\end{aligned}\quad (7)$$

as another generalisation of Eq. (3), called *maximality*. Note that $>_{\underline{P}}$ can also be viewed as a robustification of $>_{\mu}$ over μ in \mathcal{M} . Applied on the example of Section 2, we find $\text{opt}_{>_{\underline{P}}}(D) = \{1, 2, 3, 5\}$ as a solution.

Note that Walley [9, Sections 3.9.2, p. 161] has a slightly different definition: instead of working from the set of admissible decisions as in Eq. (7), Walley starts with ranking $d > e$ if $\underline{\mathbf{E}}_{\underline{P}}(J_d - J_e) > 0$ or $(J_d \geq J_e \text{ and } J_d \neq J_e)$, and then selects those decisions from D that are maximal with respect to this strict partial order. Using Theorem 3 from the appendix, it is easy to show that Walley’s definition of maximality coincides with the one given in Eq. (7) whenever the set $\{J_d : d \in D\}$ is weakly compact. This is something we usually assume to ensure the existence of admissible elements; in particular, weak compactness is assumed in Theorem 5 (see appendix). The benefit of Eq. (7) over Walley’s definition is that Eq. (7) is easier to manage in the proofs in the appendix.

3.3 Interval Dominance

Another robustification of $>_{\mu}$ is the strict partial ordering $\sqsubset_{\underline{P}}$ defined by $d \sqsubset_{\underline{P}} e$ whenever $\underline{\mathbf{E}}_{\underline{P}}(J_d) > \overline{\mathbf{E}}_{\underline{P}}(J_e)$; this means that the interval $[\underline{\mathbf{E}}_{\underline{P}}(J_d), \overline{\mathbf{E}}_{\underline{P}}(J_d)]$ is completely on the right hand side of the interval $[\underline{\mathbf{E}}_{\underline{P}}(J_e), \overline{\mathbf{E}}_{\underline{P}}(J_e)]$. The above ordering is therefore called *interval dominance* (see Zaffalon, Wesnes, and Petrini [14, Section 2.3.3, pp. 68–69] for a brief discussion and references).

$$\begin{aligned}\text{opt}_{\sqsubset_{\underline{P}}}(D) &:= \max_{\sqsubset_{\underline{P}}}(\text{opt}_{\geq}(D)) \\ &= \{d \in \text{opt}_{\geq}(D) : (\forall e \in \text{opt}_{\geq}(D))(\underline{\mathbf{E}}_{\underline{P}}(J_e) \leq \overline{\mathbf{E}}_{\underline{P}}(J_d))\}\end{aligned}\quad (8)$$

The resulting notion is weaker than maximality: applied on the example of Section 2, $\text{opt}_{\sqsubset_{\underline{P}}}(D) = \{1, 2, 3, 5, 6\}$, which is strictly larger than $\text{opt}_{>_{\underline{P}}}(D)$.

3.4 *E-Admissibility*

In the example of Section 2, we have shown that $\text{opt}_{\mathbf{E}_\mu}(D)$ may not be very robust against changes in μ . Robustifying $\text{opt}_{\mathbf{E}_\mu}(D)$ against changes of μ in \mathcal{M} , we arrive at

$$\text{opt}_{\mathcal{M}}(D) := \bigcup_{\mu \in \mathcal{M}} \text{opt}_{\mathbf{E}_\mu}(D); \quad (9)$$

this provides another way to generalise Eq. (3). The above criterion selects those admissible decisions in D that maximize expected utility with respect to at least one μ in \mathcal{M} ; i.e., they select the *E-admissible* (see Good [15, p. 114, ll. 8–9], or Levi [5, Section 4.8, p. 96, ll. 8–20]) decisions among the admissible ones. We find $\text{opt}_{\mathcal{M}}(D) = \{1, 2, 3\}$ for the example.

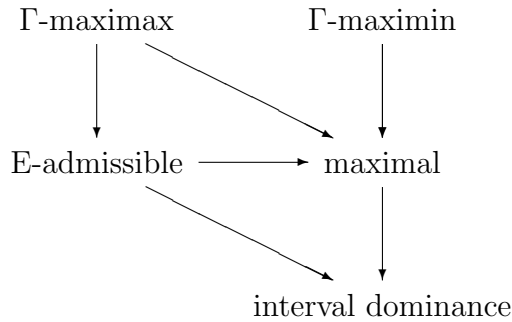
In case μ is defined on $\wp(\mathcal{X})$ and $\mu(\{x\}) > 0$ for all $x \in \mathcal{X}$, then every E-admissible decision is also admissible, and hence, in that case, $\text{opt}_{\mathcal{M}}(D)$ gives us exactly the set of E-admissible options.

4 Which Is the Right One?

Evidently, it is hard to pinpoint the right choice. Instead, let us ask ourselves: what properties do we want our notion of optimality to satisfy? Let us summarise a few important guidelines.

Clearly, whatever notion of optimality, it seems reasonable to exclude inadmissible decisions. For ease of exposition, let's assume that the inadmissible decisions have already been removed from D , i.e., $D = \text{opt}_{\geq}(D)$; this implies in particular that $\text{opt}_{\mathcal{M}}(D)$ gives us the set of E-admissible decisions.

Now note that, in general, the following implications hold:



as is also demonstrated by our example. A proof is given in the appendix, Theorem 1.

E-admissibility, maximality, and interval dominance have the nice property that the more determinate our beliefs (i.e., the smaller \mathcal{M}), the smaller the set of optimal decisions. In contradistinction, Γ -maximin and Γ -maximax lack this property, and usually only select a single decision, even in case of complete ignorance. However, if we are *only* interested in the most pessimistic (or most optimistic) solution, disregarding other reasonable solutions, then Γ -maximin (or Γ -maximax) seems appropriate. Utkin and Augustin [16] have collected a number of nice algorithms for finding Γ -maximin and Γ -maximax solutions, and even mixtures of these two. Seidenfeld [17] has compared Γ -maximin to E-admissibility, and argued against Γ -maximin in sequential decision problems.

If we do not settle for Γ -maximin (or Γ -maximax), should we choose E-admissibility, maximality, or interval dominance? As already mentioned, interval dominance is weaker than maximality, so in general we will end up with a larger (and arguably too large) set of optimal options. Assuming the non-admissible decisions have been weeded, a decision d is not optimal in D with respect to interval dominance if and only if

$$\overline{\mathbf{E}}_{\underline{P}}(J_d) < \sup_{e \in D} \underline{\mathbf{E}}_{\underline{P}}(J_e). \quad (10)$$

Thus, if D has n elements, interval dominance requires us to calculate $2n$ natural extensions, and make $2n$ comparisons, whereas for maximality, by Eq. (7), we must calculate $n^2 - n$ natural extensions, and perform $n^2 - n$ comparisons—roughly speaking, each natural extension is a linear program in m (size of \mathcal{X}) variables and r (size of \mathcal{K}) constraints, or vice versa if we solve the dual program. So, comparing maximality and interval dominance, we face a tradeoff between computational speed and number of optimal options.

However, this also means that interval dominance is a means to speed up the calculation of maximal and E-admissible decisions: because every maximal decision is also interval dominant, we can invoke interval dominance as a first computationally efficient step in eliminating non-optimal decisions, if we eventually opt for maximality or E-admissibility. Indeed, eliminating those decisions d that satisfy Eq. (10), we will also eliminate those decisions that are neither maximal, nor E-admissible.

Regarding sequential decision problems, we note that dynamic programming techniques cannot be used when using interval dominance (see De Cooman and Troffaes [18]), and therefore, since dynamic programming yields an exponential speedup, maximality and E-admissibility are certainly preferred over interval dominance

once dynamics enter the picture.

This leaves E-admissibility and maximality. They are quite similar: they coincide on all decision sets D that contain two decisions. In case we consider larger decision sets, they coincide if the set of gain gambles is convex (for instance, if we consider *randomised* decisions). As already mentioned, E-admissibility is stronger than maximality, and also has some other advantages over maximality. For instance, $\frac{1}{5}J_2 + \frac{4}{5}J_3 >_{\underline{P}} J_5$, so, choosing decision 2 with probability 20% and decision 3 with probability 80% is preferred to decision 5. Therefore, we should perhaps not consider decision 5 as optimal.

E-admissibility is not vulnerable to such argument, since no E-admissible decision can be dominated by randomized decisions: if for some $\mu \in \mathcal{M}$ it holds that $\mathbf{E}_\mu(J_d - J_e) \geq 0$ for all $e \in D$, then also

$$\mathbf{E}_\mu \left(J_d - \sum_{i=1}^n \lambda_i J_{e_i} \right) = \sum_{i=1}^n \lambda_i \mathbf{E}_\mu (J_d - J_{e_i}) \geq 0$$

for any convex combination $\sum_{i=1}^n \lambda_i J_{e_i}$ of gain gambles, and hence, it also holds that $\underline{\mathbf{E}}_{\underline{P}} (\sum_{i=1}^n \lambda_i J_{e_i} - J_d) \leq 0$ which means that no convex combination $\sum_{i=1}^n \lambda_i J_{e_i}$ can dominate J_d with respect to $>_{\underline{P}}$.

A powerful algorithm for calculating E-admissible options has been recently suggested by Utkin and Augustin [16, pp. 356–357], and independently by Kikuti, Cozman, and de Campos [19, Sec. 3.4]. If D has n elements, finding all (pure) E-admissible options requires us to solve n linear programs in m variables and $r + n$ constraints.

As we already noted, through convexification of the decision set, maximality and E-admissibility coincide. Utkin and Augustin’s algorithm can also cope with this case, but now one has to consider in the worst case $n!$ linear programs, and usually several less: the worst case only obtains if all options are E-admissible. For instance, if there are only ℓ E-admissible pure options, one has to consider only at most $\ell! + n - \ell$ of those linear programs, and again, usually less.

In conclusion, the decision criterion to settle for in a particular application, depends at least on the goals of the decision maker (what properties should optimality satisfy?), and possibly also on the size and structure of the problem if computational issues arise.

Acknowledgements

I especially want to thank Teddy Seidenfeld for the many instructive discussions about maximality versus E-admissibility. I also wish to thank two anonymous referees for their helpful comments. This paper has been supported by the Belgian American Educational Foundation. The scientific responsibility rests with its author.

A Proofs

This appendix is dedicated to proving the connections between the various optimality criteria, and existence results mentioned throughout the paper. In the whole appendix, we assume the following:

Recall, D denotes some set of decisions, and every decision $d \in D$ induces a gain gamble $J_d \in \mathcal{L}(\mathcal{X})$, where $\mathcal{L}(\mathcal{X})$ is the set of all gambles (bounded \mathcal{X} - \mathbb{R} mappings).

\underline{P} denotes a lower prevision, defined on a subset \mathcal{K} of $\mathcal{L}(\mathcal{X})$. With \mathcal{F} we denote a field on \mathcal{X} such that all gain gambles J_d and gambles in \mathcal{K} are measurable with respect to \mathcal{F} , i.e., are a uniform limit of \mathcal{F} -simple gambles. \mathcal{F} could be for instance the power set of \mathcal{X} .

\underline{P} is assumed to avoid sure loss, and $\underline{\mathbf{E}}_{\underline{P}}$ is its natural extension to the set of all \mathcal{F} -measurable gambles. \mathcal{M} is the credal set representing \underline{P} , as defined in Section 3. We will make deliberate use of the properties of natural extension (for instance, superadditivity: $\underline{\mathbf{E}}_{\underline{P}}(f + g) \geq \underline{\mathbf{E}}_{\underline{P}}(f) + \underline{\mathbf{E}}_{\underline{P}}(g)$, and hence also $\underline{\mathbf{E}}_{\underline{P}}(f - g) \leq \underline{\mathbf{E}}_{\underline{P}}(f) - \underline{\mathbf{E}}_{\underline{P}}(g)$). We refer to Walley [9, Sec. 2.6, p. 76, and Sec. 3.1.2, p. 123] for an overview and proof of these properties.

We use the symbol μ for an arbitrary finitely additive probability measure on \mathcal{F} , and \mathbf{E}_{μ} denotes the Dunford integral with respect to μ . This integral is defined on (at least) the set of all \mathcal{F} -measurable gambles.

A.1 Connections between Decision Criteria

Theorem 1 *The following relations hold.*

$$\begin{aligned} \text{opt}_{\overline{\mathbf{E}}_P}(D) &\subseteq \text{opt}_{\mathcal{M}}(D) \subseteq \text{opt}_{>_P}(D) \subseteq \text{opt}_{\sqsupseteq_P}(D) \\ \text{opt}_{\underline{\mathbf{E}}_P}(D) &\subseteq \text{opt}_{>_P}(D) \end{aligned}$$

PROOF. Let $\mathcal{J} = \{J_d : d \in D\}$.

Suppose that d is Γ -maximax in D : J_d maximises $\overline{\mathbf{E}}_P$ in $\max_{\geq}(\mathcal{J})$. Since $\overline{\mathbf{E}}_P$ is the upper envelope of \mathcal{M} , and \mathcal{M} is weak-* compact (see Walley [9, Sec. 3.6]), there is a μ in \mathcal{M} such that $\overline{\mathbf{E}}_P(J_d) = \mathbf{E}_{\mu}(J_d)$. But, $\mathbf{E}_{\mu}(J_e) \leq \overline{\mathbf{E}}_P(J_e) \leq \overline{\mathbf{E}}_P(J_d) = \mathbf{E}_{\mu}(J_d)$, for every $J_e \in \max_{\geq}(\mathcal{J})$ because d is Γ -maximax. Thus, d belongs to $\text{opt}_{\mathcal{M}}(D)$.

Suppose that $d \in \text{opt}_{\mathcal{M}}(D)$: there is a μ in \mathcal{M} such that J_d maximises \mathbf{E}_{μ} in $\max_{\geq}(\mathcal{J})$. But then, because $\underline{\mathbf{E}}_P$ is the lower envelope of \mathcal{M} , $\underline{\mathbf{E}}_P(J_e - J_d) \leq \mathbf{E}_{\mu}(J_e - J_d) = \mathbf{E}_{\mu}(J_e) - \mathbf{E}_{\mu}(J_d) \leq 0$ for all J_e in $\max_{\geq}(\mathcal{J})$. Hence, by Eq. (7) on p. 8, d must be maximal.

Suppose that d is maximal. Then, again by Eq. (7), $\underline{\mathbf{E}}_P(J_e - J_d) \leq 0$ for all J_e in $\max_{\geq}(\mathcal{J})$. But, $\underline{\mathbf{E}}_P(J_e) - \overline{\mathbf{E}}_P(J_d) \leq \underline{\mathbf{E}}_P(J_e - J_d)$, hence, also $\underline{\mathbf{E}}_P(J_e) \leq \overline{\mathbf{E}}_P(J_d)$ for all J_e in $\max_{\geq}(\mathcal{J})$, which means that d belongs to $\text{opt}_{\sqsupseteq_P}(D)$.

Finally, suppose that d is Γ -maximin: J_d maximises $\underline{\mathbf{E}}_P$ in $\max_{\geq}(\mathcal{J})$. But then $\underline{\mathbf{E}}_P(J_e - J_d) \leq \underline{\mathbf{E}}_P(J_e) - \underline{\mathbf{E}}_P(J_d) \leq 0$ for all J_e in $\max_{\geq}(\mathcal{J})$; d must be maximal.

A.2 Existence

We first prove a technical but very useful lemma about the existence of optimal elements with respect to preorders; it's an abstraction of a result proved by De Cooman and Troffaes [18]. Let's start with a few definitions.

A *preorder* is simply a reflexive and transitive relation.

Let \mathcal{V} be any set, and let \succsim be any preorder on \mathcal{V} . An element v of a subset \mathcal{S} of \mathcal{V} is called *\succsim -maximal* in \mathcal{S} if, for all w in \mathcal{S} , $w \succsim v$ implies $v \succsim w$. The set of

\triangleright -maximal elements is denoted by

$$\max_{\triangleright}(\mathcal{S}) := \left\{ v \in \mathcal{S} : (\forall w \in \mathcal{S})(w \triangleright v \implies v \triangleright w) \right\}. \quad (\text{A.1})$$

For any v in \mathcal{S} , we also define the *up-set* of v relative to \mathcal{S} as

$$\uparrow_{\triangleright}^{\mathcal{S}} v := \{ w \in \mathcal{S} : w \triangleright v \}.$$

Lemma 2 *Let \mathcal{V} be a Hausdorff topological space. Let \triangleright be any preorder on \mathcal{V} such that for any v in \mathcal{V} , the set $\uparrow_{\triangleright}^{\mathcal{V}} v$ is closed. Then, for any non-empty compact subset \mathcal{S} of \mathcal{V} , the following statements hold.*

- (i) *For every v in \mathcal{S} , the set $\uparrow_{\triangleright}^{\mathcal{S}} v$ is non-empty and compact.*
- (ii) *The set $\max_{\triangleright}(\mathcal{S})$ of \triangleright -maximal elements of \mathcal{S} is non-empty.*
- (iii) *For every v in \mathcal{S} , there is a \triangleright -maximal element w of \mathcal{S} such that $w \triangleright v$.*

PROOF. (i). Since \triangleright is reflexive, it follows that $v \triangleright v$, so $\uparrow_{\triangleright}^{\mathcal{S}} v$ is non-empty. Is it compact? Clearly, $\uparrow_{\triangleright}^{\mathcal{S}} v = \mathcal{S} \cap \uparrow_{\triangleright}^{\mathcal{V}} v$, so $\uparrow_{\triangleright}^{\mathcal{S}} v$ is the intersection of a compact set and a closed set, and therefore $\uparrow_{\triangleright}^{\mathcal{S}} v$ must be compact too.

(ii). Let \mathcal{S}' be any subset of the non-empty compact set \mathcal{S} that is linearly ordered with respect to \triangleright . If we can show that \mathcal{S}' has an upper bound in \mathcal{S} with respect to \triangleright , then we can infer from a version of Zorn's lemma [20, (AC7), p. 144] (which also holds for preorders) that \mathcal{S} has a \triangleright -maximal element. Let then $\{v_1, v_2, \dots, v_n\}$ be an arbitrary finite subset of \mathcal{S}' . We can assume without loss of generality that $v_1 \triangleright v_2 \triangleright \dots \triangleright v_n$, and consequently $\uparrow_{\triangleright}^{\mathcal{S}} v_1 \subseteq \uparrow_{\triangleright}^{\mathcal{S}} v_2 \subseteq \dots \subseteq \uparrow_{\triangleright}^{\mathcal{S}} v_n$. This implies that the intersection $\bigcap_{k=1}^n \uparrow_{\triangleright}^{\mathcal{S}} v_k = \uparrow_{\triangleright}^{\mathcal{S}} v_1$ of these up-sets is non-empty: the collection $\{\uparrow_{\triangleright}^{\mathcal{S}} v : v \in \mathcal{S}'\}$ of compact and hence closed (\mathcal{V} is Hausdorff) subsets of \mathcal{S} has the finite intersection property. Consequently, since \mathcal{S} is compact, the intersection $\bigcap_{v \in \mathcal{S}'} \uparrow_{\triangleright}^{\mathcal{S}} v$ is non-empty as well, and this is the set of upper bounds of \mathcal{S}' in \mathcal{S} with respect to \triangleright . So, by Zorn's lemma, \mathcal{S} has a \triangleright -maximal element: $\max_{\triangleright}(\mathcal{S})$ is non-empty.

(iii). Combine (i) and (ii) to show that the non-empty compact set $\uparrow_{\triangleright}^{\mathcal{S}} v$ has a maximal element w with respect to \triangleright . It is then a trivial step to prove that w is also \triangleright -maximal in \mathcal{S} : we must show that for any u in \mathcal{S} , if $u \triangleright w$, then $w \triangleright u$. But, if $u \triangleright w$, then also $u \triangleright v$ since $w \triangleright v$ by construction. Hence, $u \in \uparrow_{\triangleright}^{\mathcal{S}} v$, and since w is \triangleright -maximal in $\uparrow_{\triangleright}^{\mathcal{S}} v$, it follows that $w \triangleright u$.

The weak topology on $\mathcal{L}(\mathcal{X})$ is simply the topology of point-wise convergence.

That is, a net f_α in $\mathcal{L}(\mathcal{X})$ converges weakly to f in $\mathcal{L}(\mathcal{X})$ if $\lim_\alpha f_\alpha(x) = f(x)$ for all $x \in \mathcal{X}$.

Theorem 3 *If $\mathcal{J} = \{J_d: d \in D\}$ is a non-empty and weakly compact set, then D contains at least one admissible decision, and even more, for every decision e in D , there is an admissible decision d in D such that $J_d \geq J_e$.*

PROOF. It is easy to derive from Eq. (2) that

$$\text{opt}_{\geq}(D) = \{d \in D: (\forall e \in D)(J_e \geq J_d \implies J_d \geq J_e)\}.$$

Hence, a decision is admissible in D exactly when its gain gamble is \geq -maximal in \mathcal{J} . We must show that \mathcal{J} has \geq -maximal elements.

By Lemma 2, it suffices to prove that, for every $f \in \mathcal{L}(\mathcal{X})$, the set $\mathcal{G}_f = \{g \in \mathcal{L}(\mathcal{X}): g \geq f\}$ is closed with respect to the topology of point-wise convergence.

Let g_α be a net in \mathcal{G}_f , and suppose that g_α converges point-wise to $g \in \mathcal{L}(\mathcal{X})$: for every $x \in \mathcal{X}$, $\lim_\alpha g_\alpha(x) = g(x)$. But, since $g_\alpha(x) \geq f(x)$ for every \mathcal{X} , it must also hold that $g(x) = \lim_\alpha g_\alpha(x) \geq f(x)$. Hence, $g \in \mathcal{G}_f$. We have shown that every converging net in \mathcal{G}_f converges to a point in \mathcal{G}_f . Thus, \mathcal{G}_f is closed. This establishes the theorem.

Let's now introduce a slightly stronger topology on $\mathcal{L}(\mathcal{X})$. This topology has no particular name in the literature, so let's just call it the τ -topology. It is determined by the following convergence.

Definition 4 *Say that a net f_α in $\mathcal{L}(\mathcal{X})$ τ -converges to f in $\mathcal{L}(\mathcal{X})$, if*

- (i) $\lim_\alpha f_\alpha(x) = f(x)$ for all $x \in \mathcal{X}$ (point-wise convergence), and
- (ii) $\lim_\alpha \bar{\mathbf{E}}_{\underline{P}}(|f_\alpha - f|) = 0$ (convergence in $\bar{\mathbf{E}}_{\underline{P}}(|\cdot|)$ -norm).

This convergence induces a topology τ on $\mathcal{L}(\mathcal{X})$: it turns $\mathcal{L}(\mathcal{X})$ into a locally convex topological vector space, which also happens to be Hausdorff. A topological basis at 0 consists for instance of the convex sets

$$\{f \in \mathcal{L}(\mathcal{X}): \bar{\mathbf{E}}_{\underline{P}}(|f|) < \epsilon \text{ and } f(x) < \delta(x)\},$$

for $\epsilon > 0$, and $\delta(x) > 0$ for all $x \in \mathcal{X}$. It has more open sets and more closed sets than the weak topology, but it has less compact sets than the weak topology. On the other hand, this topology is weaker than the supremum norm topology,

so it has fewer open and closed sets, and more compact sets, compared to the supremum norm topology. Note that in case \mathcal{X} is finite, it reduces to the weak topology, which is in that case also equivalent to the supremum norm topology.

Note that $\underline{\mathbf{E}}_P$, $\overline{\mathbf{E}}_P$, and \mathbf{E}_μ for all $\mu \in \mathcal{M}$, are τ -continuous, simply because

$$\begin{aligned}\overline{\mathbf{E}}_P(|f_\alpha - f|) &\geq |\underline{\mathbf{E}}_P(f_\alpha) - \underline{\mathbf{E}}_P(f)|, \\ \overline{\mathbf{E}}_P(|f_\alpha - f|) &\geq |\overline{\mathbf{E}}_P(f_\alpha) - \overline{\mathbf{E}}_P(f)|, \text{ and} \\ \overline{\mathbf{E}}_P(|f_\alpha - f|) &\geq |\mathbf{E}_\mu(f_\alpha) - \mathbf{E}_\mu(f)|\end{aligned}$$

(see Walley [9, p. 77, Sec. 2.6.1(l)]). We will exploit this fact in the proof of the following theorem, generalising a result due to Walley [9, p. 161, Sec. 3.9.2].

Theorem 5 *If $\mathcal{J} = \{J_d : d \in D\}$ is non-empty and compact with respect to the τ -topology, then the following statements hold.*

- (i) $\text{opt}_{\mathbf{E}_\mu}(D)$ is non-empty for all $\mu \in \mathcal{M}$.
- (ii) $\text{opt}_{\underline{\mathbf{E}}_P}(D)$ is non-empty.
- (iii) $\text{opt}_{\overline{\mathbf{E}}_P}(D)$ is non-empty.
- (iv) $\text{opt}_{>_P}(D)$ is non-empty.
- (v) $\text{opt}_{\sqsupseteq_P}(D)$ is non-empty.
- (vi) $\text{opt}_{\mathcal{M}}(D)$ is non-empty.

PROOF. (i). Introduce the following order on $\mathcal{L}(\mathcal{X})$: say that $f \triangleright g$ whenever $\mathbf{E}_\mu(f) \geq \mathbf{E}_\mu(g)$. Let's first show that, for all $f \in \mathcal{L}(\mathcal{X})$, the set $\mathcal{G}_f = \{g \in \mathcal{L}(\mathcal{X}) : g \triangleright f\}$ is τ -closed.

Let g_α be a net in \mathcal{G}_f , and suppose that g_α τ -converges to $g \in \mathcal{L}(\mathcal{X})$. Since the integral \mathbf{E}_μ is τ -continuous, it follows that $\mathbf{E}_\mu(g) = \lim_\alpha \mathbf{E}_\mu(g_\alpha) \geq \mathbf{E}_\mu(f)$. Concluding, g belongs to \mathcal{G}_f . We have established that every converging net in \mathcal{G}_f converges to a point in \mathcal{G}_f . Thus, \mathcal{G}_f is τ -closed.

By Lemma 2, it follows that \mathcal{J} has at least one \triangleright -maximal element J_e , that is, J_e maximises \mathbf{E}_μ in \mathcal{J} . Since any τ -compact set is also weakly compact, there is a \geq -maximal element J_d in \mathcal{J} such that $J_d \geq J_e$, by Theorem 3. But then, $\mathbf{E}_\mu(J_d) \geq \mathbf{E}_\mu(J_e)$, and hence, J_d also maximises \mathbf{E}_μ in \mathcal{J} . Because J_d is \geq -maximal in \mathcal{J} , it also maximises \mathbf{E}_μ in $\max_{\geq}(\mathcal{J})$. This establishes that d belongs to $\text{opt}_{\mathbf{E}_\mu}(D)$: this set is non-empty.

(ii). Introduce the following order on $\mathcal{L}(\mathcal{X})$: say that $f \triangleright g$ whenever $\underline{\mathbf{E}}_P(f) \geq$

$\underline{E}_P(g)$. Continue along the lines of (i), using the fact that \underline{E}_P is τ -continuous.

(iii). Again along the lines of (i), with $f \supseteq g$ whenever $\overline{E}_P(f) \geq \overline{E}_P(g)$.

(iv)&(v)&(vi). Immediate, by (iii) and Theorem 1.

References

- [1] J. von Neumann, O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press, 1944.
- [2] I. N. Herstein, J. Milnor, An axiomatic approach to measurable utility, *Econometrica* 21 (2) (1953) 291–297.
- [3] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*, 2nd Edition, Springer, 1985.
- [4] H. Raiffa, R. Schlaifer, *Applied Statistical Decision Theory*, MIT Press, 1961.
- [5] I. Levi, *The Enterprise of Knowledge. An Essay on Knowledge, Credal Probability, and Chance*, MIT Press, Cambridge, 1983.
- [6] N. Dunford, Integration in general analysis, *Transactions of the American Mathematical Society* 37 (3) (1935) 441–453.
- [7] N. Dunford, J. T. Schwartz, *Linear Operators*, John Wiley & Sons, New York, 1957.
- [8] O. Kallenberg, *Foundations of Modern Probability*, 2nd Edition, Probability and Its Applications, Springer, 2002.
- [9] P. Walley, *Statistical Reasoning with Imprecise Probabilities*, Chapman and Hall, London, 1991.
- [10] F. J. Giron, S. Rios, Quasi-Bayesian behaviour: A more realistic approach to decision making?, in: J. M. Bernardo, J. H. DeGroot, D. V. Lindley, A. F. M. Smith (Eds.), *Bayesian Statistics*, University Press, Valencia, 1980, pp. 17–38.
- [11] B. De Finetti, *Theory of Probability: A Critical Introductory Treatment*, Wiley, New York, 1974–5, two volumes.
- [12] I. Gilboa, D. Schmeidler, Maxmin expected utility with non-unique prior, *Journal of Mathematical Economics* 18 (2) (1989) 141–153.
- [13] J. K. Satia, J. Roy E. Lave, Markovian decision processes with uncertain transition probabilities, *Operations Research* 21 (3) (1973) 728–740.

- [14] M. Zaffalon, K. Wesnes, O. Petrini, Reliable diagnoses of dementia by the naive credal classifier inferred from incomplete cognitive data, *Artificial Intelligence in Medicine* 29 (1–2) (2003) 61–79.
- [15] I. J. Good, Rational decisions, *Journal of the Royal Statistical Society, Series B* 14 (1) (1952) 107–114.
- [16] L. V. Utkin, T. Augustin, Powerful algorithms for decision making under partial prior information and general ambiguity attitudes, in: F. G. Cozman, R. Nau, T. Seidenfeld (Eds.), *Proceedings of the Fourth International Symposium on Imprecise Probabilities and Their Applications*, 2005, pp. 349–358.
- [17] T. Seidenfeld, A contrast between two decision rules for use with (convex) sets of probabilities: Gamma-maximin versus E-admissibility, *Synthese* 140 (1–2) (2004) 69–88.
- [18] G. de Cooman, M. C. M. Troffaes, Dynamic programming for deterministic discrete-time systems with uncertain gain, *International Journal of Approximate Reasoning* 39 (2–3) (2004) 257–278.
- [19] D. Kikuti, F. G. Cozman, C. P. de Campos, Partially ordered preferences in decision trees: Computing strategies with imprecision in probabilities, in: R. Brafman, U. Junker (Eds.), *Multidisciplinary IJCAI-05 Workshop on Advances in Preference Handling*, 2005, pp. 118–123.
- [20] E. Schechter, *Handbook of Analysis and Its Foundations*, Academic Press, San Diego, 1997.